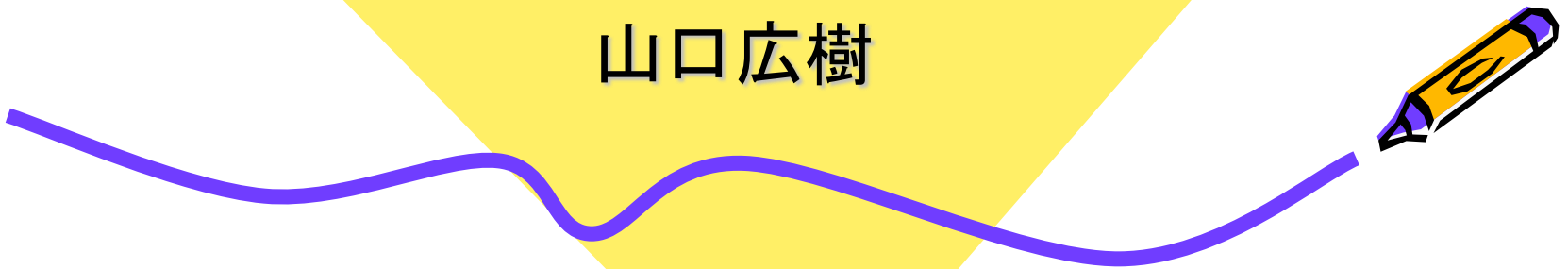


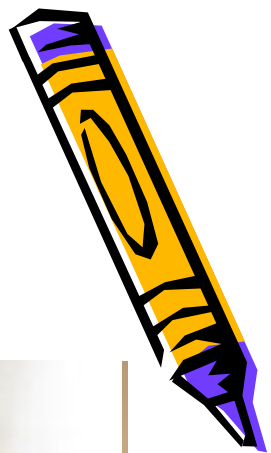


未来を拓くアホ発見器

修士課程2年
山口広樹



よくある風景



研究目的

周りに排除されてしまっている発言（あほ発言）を
抽出して再検討することで埋もれてしまったココロ
をマイニングする

真意,
感情, 私見
物語



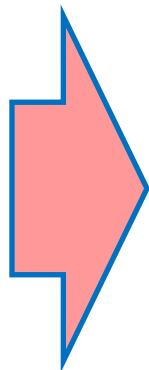
アプローチ



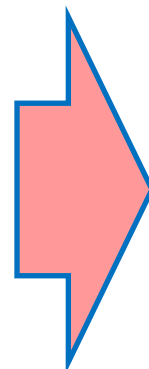
- I. 会話ログを解析対象とし、**テキストマイニング**をしてあほ発言を自動抽出する.

入力

会話
ログ



テキスト処理



出力

あほ
発言

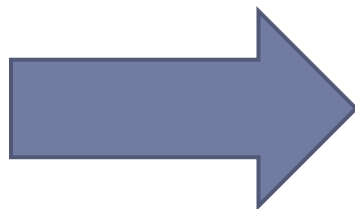


入力：会話テキスト→前処理

C 明日の朝食何にする？
B 卵かけごはんがいいね
C サンドイッチがいいな
A 男は黙って納豆定食
B サンドイッチもいいね
C やっぱお茶漬けがいい
A 納豆...

会話テキストの例

形態素解析



C,明日,の,朝食,何,に,する,?
B,卵,かけ,ごはん,が,いい,ね
C,サンドイッチ,が,いい,な
A,男,は,黙っ,て,納豆,定食
B,サンドイッチ,も,いい,ね
C,やっぱ,お茶,漬け,が,いい
A,納豆

形態素解析

発言ベクトル生成

$$\mathbf{U}_k = (t_0, t_1, \dots, t_n), |\mathbf{U}_k| = 1 \quad (t_k: k\text{番目の単語の出現頻度})$$

発言間類似度算出

$$\begin{array}{l} \text{前} \\ \text{後} \end{array} \begin{array}{l} S_{\text{before}\cdot n} = W_{\text{before}\cdot n}^T \cdot U_n \\ S_{\text{after}\cdot n} = U_n \cdot W_{\text{after}\cdot n}^T \end{array} \left(\begin{array}{l} W_{\text{before}\cdot n} = \sum_{k < n} U_k \\ W_{\text{after}\cdot n}^T = \sum_{k > n} U_k \end{array} , |W_{\text{before}\cdot n}| = |W_{\text{after}\cdot n}| = 1 \right)$$

※前後の窓付

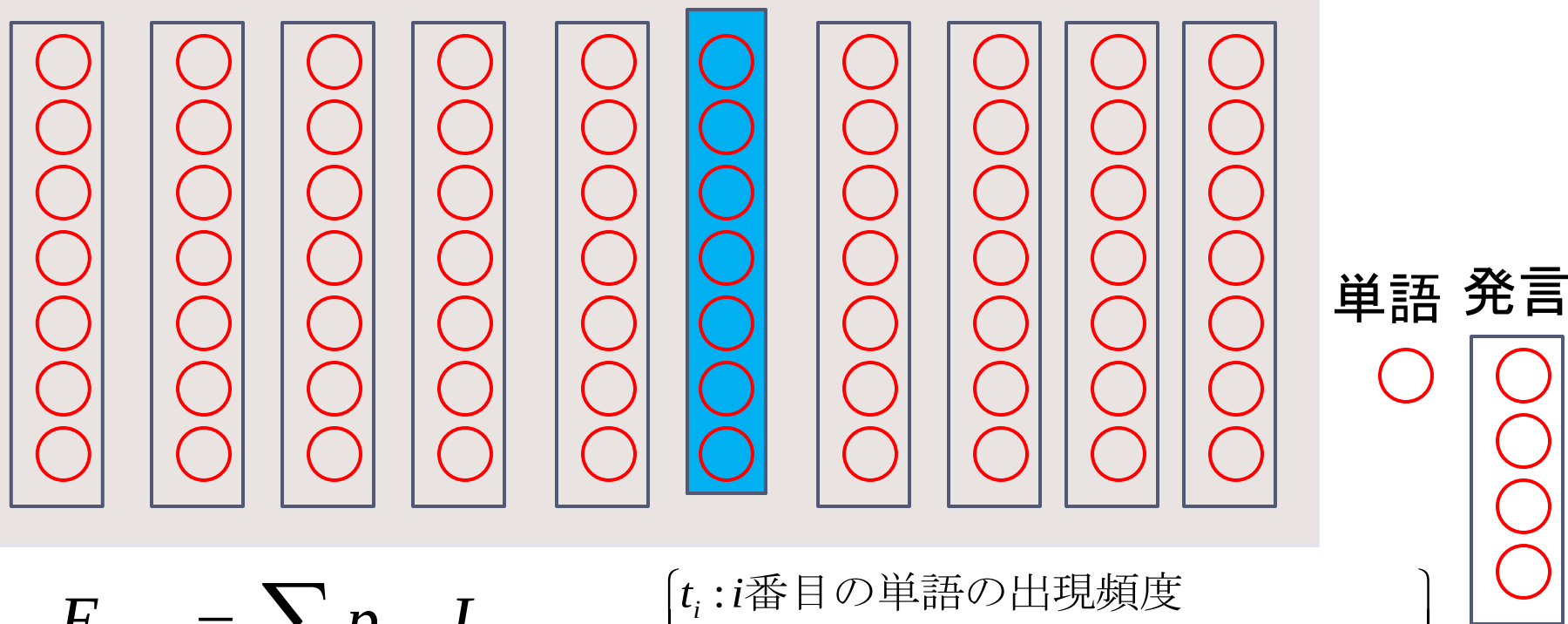
あほ発言

前後コンテキストとの関連性が低く、排除されてしまっている発言

例外発言の抽出条件

$$S_{\text{before}\cdot n} < \alpha \cap S_{\text{after}\cdot n} < \beta \quad (\alpha, \beta \text{は閾値})$$

あほ発言の性質評価-選択平均情報量-



$$E_{\mathbf{AU}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{A} t_i}$$

但し、

$$I_{\mathbf{A} t_i} = -\log p_{\mathbf{A} t_i}$$

t_i : i 番目の単語の出現頻度
 $\mathbf{U}_k$: 注目発言ベクトル
 $p_{\mathbf{U}_k t_i}$: $\mathbf{U}_k$ における単語 t_i の生起確率
 $p_{\mathbf{A} t_i}$: 集合 $\mathbf{A}$ における単語 t_i の生起確率

あほ発言の性質評価-選択平均情報量-

$$E_{\mathbf{A} \mathbf{U}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{A} t_i} \quad \text{但し, } \left\{ \begin{array}{l} t_i : i\text{番目の単語の出現頻度} \\ \mathbf{U}_k : \text{注目発言ベクトル} \\ p_{\mathbf{U}_k t_i} : \mathbf{U}_k \text{における単語 } t_i \text{の生起確率} \\ p_{\mathbf{A} t_i} : \text{集合 } \mathbf{A} \text{における単語 } t_i \text{の生起確率} \end{array} \right.$$

$$I_{\mathbf{A} t_i} = -\log p_{\mathbf{A} t_i}$$

$$E_{\mathbf{W}_{\text{before } k} \mathbf{U}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{W}_{\text{before } k} t_i}$$

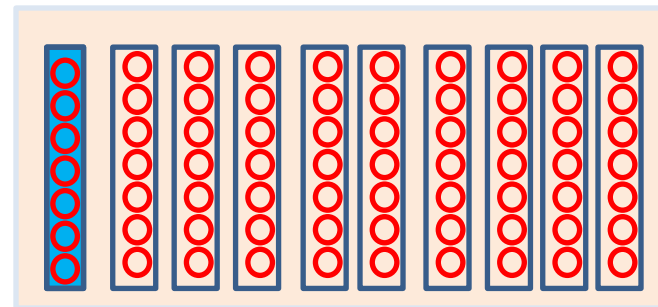
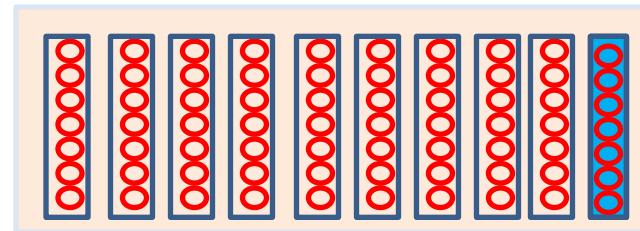
$$E_{\mathbf{W}_{\text{after } k} \mathbf{U}_k} = \sum_{t_i \in \mathbf{U}_k} p_{\mathbf{U}_k t_i} I_{\mathbf{W}_{\text{after } k} t_i}$$

$$\left(\begin{array}{l} W_{\text{before } n} = \sum_{k \leq n} \mathbf{U}_k \\ W_{\text{after } n}^T = \sum_{k \leq n} \mathbf{U}_k \end{array} \right), \quad |W_{\text{before } n}| = |W_{\text{after } n}| = 1$$

※前後の窓付

評価関数

$$\bar{E} = E_{\mathbf{W}_{\text{before } k} \mathbf{U}_k} - E_{\mathbf{W}_{\text{after } k} \mathbf{U}_k}$$



あほ発言の性質評価-選択平均情報量-

$$\bar{E} = E_{w_{beforek} U_k} - E_{w_{after \cdot k} U_k}$$

$$\bar{E} \text{ 大} \Rightarrow E_{w_{beforek} U_k} \text{ 大} \text{ 且 } E_{w_{after \cdot k} U_k} \text{ 小}$$

新しい話題を導入

$$\bar{E} \text{ 小} \Rightarrow E_{w_{beforek} U_k} \doteq E_{w_{after \cdot k} U_k}$$

会話コンテキストに特に影響なし

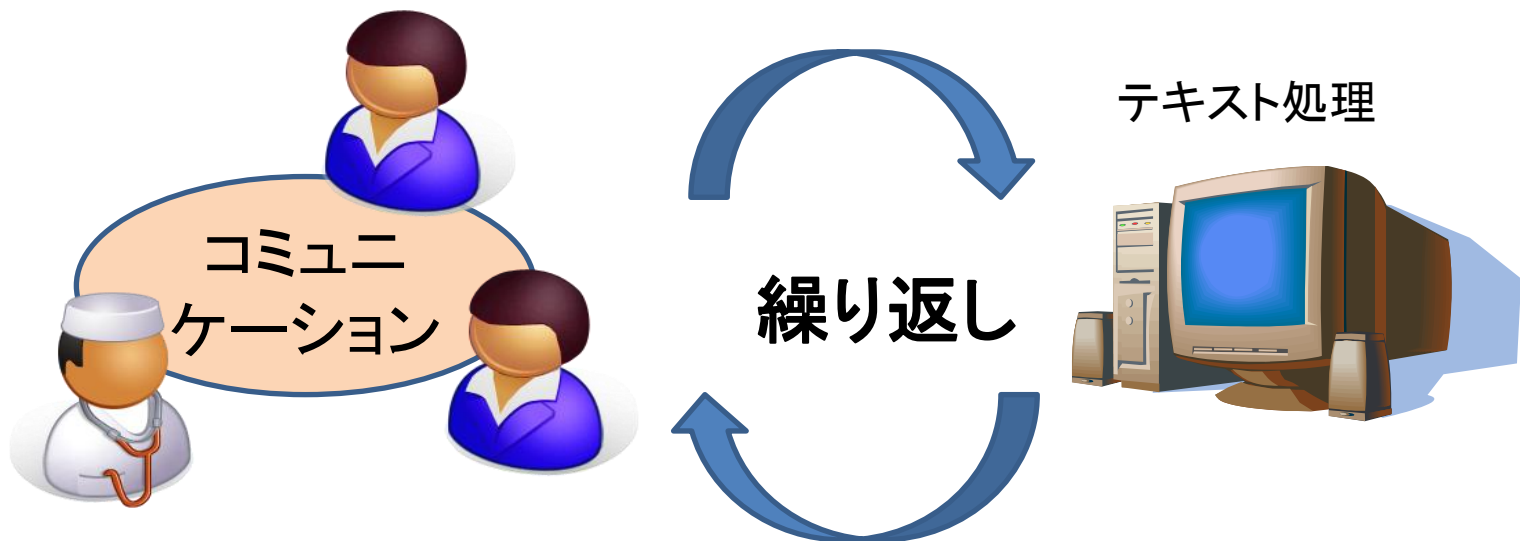
出力:あほ発言

$\bar{E}$ をスコアとしてランク付して出力

Rank	発言No	発言者	発言内容	Score
------	------	-----	------	-------

提案システム利用法

- I. 会話ログをテキスト処理によって解析し、
あほ発言を自動抽出.
- II. 抽出した発言を基に再検討(新たな会話が
が発生し、ログデータ入手).
- III. I.とII.を繰り返し行う.



オンラインチャット実験

実験概要

テーマ『10分間で楽しめる新しいゲームを考えよう』

被験者 6人

30分～40分程度
全128発言

1回目の解析結果をフィードバックした後、
2回目の実験を行った。

オンラインチャット実験結果

No	順	名	発言内容	点
66	1	C	あ、身体使う系だ。新しい＞F	0.301
72	4	A	やっぱみんなで遊ぼうよ	0.190
32	6	F	じゃあ僕は6で	0.187
103	7	E	んで、間違ったらサドンデス！！＞C	0.187
84	10	B	リアル鬼ごっこ. 90人で10人を追いかける	0.157
51	13	A	マスには乾杯とかばかり	0.139
105	18	C	ジャッジがいるよな。＞E	0.121



オンラインチャット実験結果

2回目のチャット実験で利用された例外発言

C	あ、身体使う系だ。新しい
A	やっぱみんなで遊ぼうよ
E	んで、間違ったらサドンデス！！＞C
C	ジャッジがいるよな。＞E

まとめ

- 本研究によって、従来では潜在的価値を発揮することのできなかった発言を抽出し、その価値をマイニングすることの有効性を示した。
- あほ発言には、通常見落とされがちな意見・主張(ココロ)が含まれている可能性があることを示した。

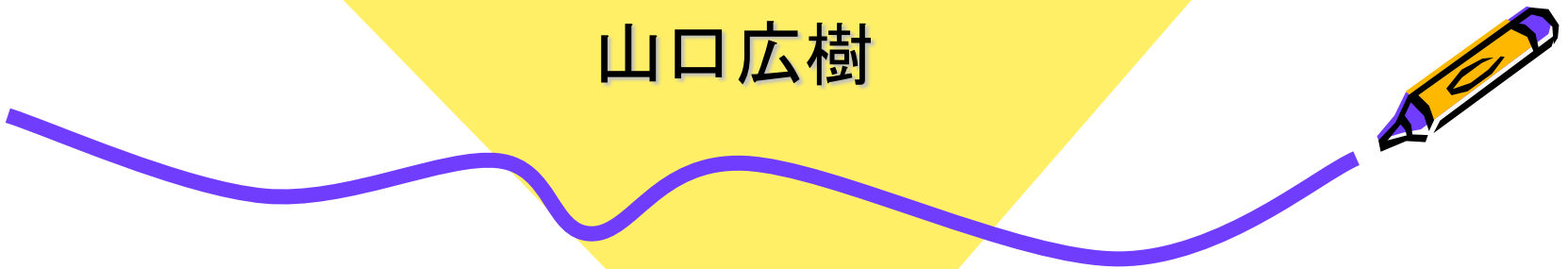
今後の展望

- 本手法を用いて創造した新アイデアの評価手法を確立する。
- 発言の性質評価から、会話を活性化させるような発言を自動生成するシステムを開発する。
- 発言者・発言の方向を考慮できるようアルゴリズムを改良する。
- 本手法の他分野における応用可能性を模索する。



未来を拓くアホ発見器

修士課程2年
山口広樹





繰り返し

テキスト処理

